



Partial Cumulative Incremental Backups Offer Unique Advantages to Customers

Teradactyl LLC.
2301 Yale Blvd. S.E., Suite C-7
Albuquerque, NM 87106
U.S. & International Sales: (505) 242-1091

info@teradactyl.com

www.teradactyl.com

Introduction

The True incremental Backup System (TiBS) now supports a new backup technique, the Partial Cumulative Incremental Backup (PCI Backup). This paper provides an overview of how PCI Backups work and how they can be used as a replacement for Cumulative Incremental Backups (CI Backups). Typically much smaller than currently available CI Backups, PCI Backups extend the patented TeraMerge multiple level synthetic backup consolidation process to include multiple backup volumes at each level of consolidation. These new, disk based PCI Backups allow for increased near line storage of backups, faster synthetic backup consolidation, improved efficiencies in backup hardware, and faster average restore times.

Comparison of PCI Backups and CI Backups

One way to understand what a PCI Backup is and how it behaves is to compare it to the more traditional CI Backup. The CI Backup is commonly used by vendors that support multiple backup levels. These multiple level backups are usually defined as level 0 through level 9. A level 0 backup is usually considered a “full” or “complete” backup. Many complex level backup strategies may be implemented. Multiple levels of backup allow site administrators to reduce the frequency of periodic full backups. For simplicity, we will consider a typical 3 level strategy consisting of monthly (level 0), weekly (level 1) and daily (level 2) backups.

How Cumulative Incremental Backups Behave

A Cumulative Incremental Backup will backup the files that have been modified or created *since the most recent lower level backup* (level n-1 or lower). For example, a special type of backup known as a Differential Backup (level 1) copies all of the files that have been modified or created since a previous Full Backup (level 0).

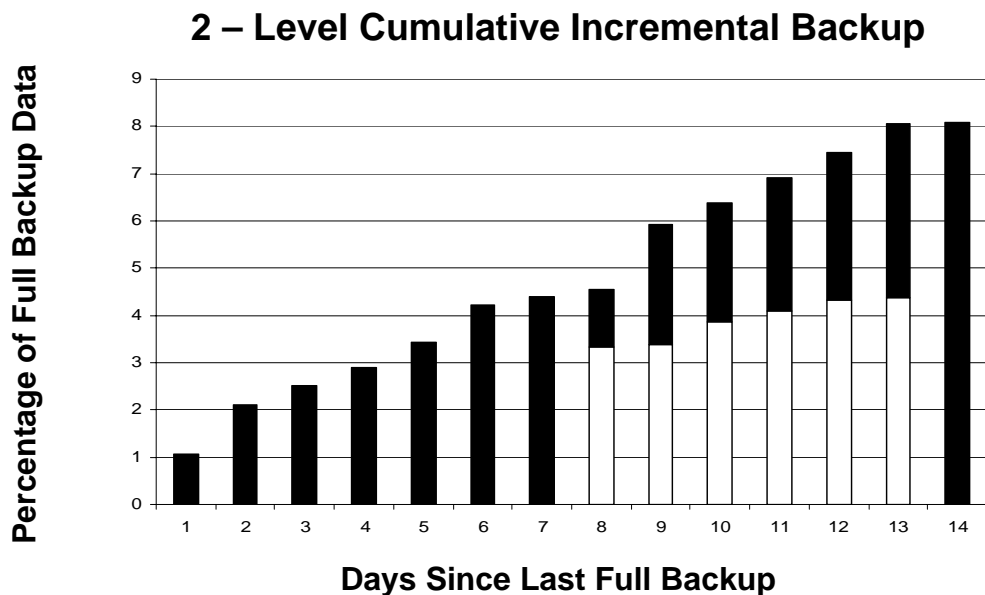


Figure 1

Figure 1 depicts how the sizes of two cumulative incremental backup levels, daily (level 2) and weekly (level 1), can change over time. The shape of the curve represents the behavior of a Differential Backup taken over the same time period. Thus, the white space below Days 8 through 13 represents the reduction in the backup size that a three level Cumulative Incremental Backup provides over a Differential Backup strategy. The level 2 daily backup will continue to grow each day as more data is changed on the backup client. Once the level 1 weekly backup completes (Days 7 and 14), the size of the level 2 daily backup is reset and the process starts over. Each week, the size of the level 1 weekly backup will continue to grow until a new level 0 full backup (not shown) is generated. The new full backup will reset the size of the weekly and daily backups and a new backup cycle is started. We use the term cumulative to describe the way this type of backup *accumulates data* at any given backup level over time.

How Partial Cumulative Incremental Backups Behave

A Partial Cumulative Incremental Backup is a specialized form of Incremental Backup which backs up the files that have been modified or created *since the most recent backup at the same level or lower* (level n or lower). For example, a True incremental Backup will only backup the files that have been modified or created since the most recent backup from the client (at any backup level).

2 – Level Partial Cumulative Incremental Backup

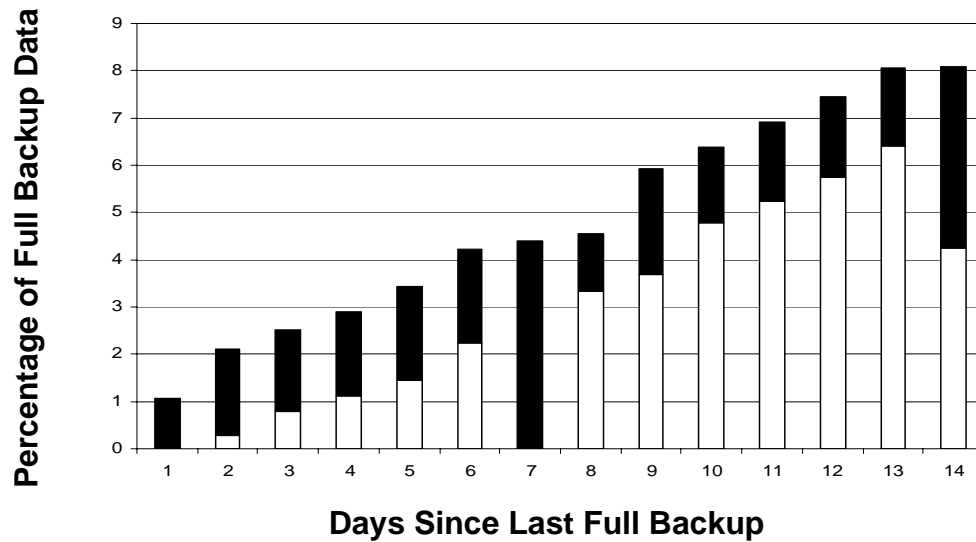


Figure 2

Figure 2 shows how the sizes of two partial cumulative incremental backup levels, daily (level 2) and weekly (level 1), can change over time. The level 2 daily backup copies only changes since the last client backup, essentially flattening the size of each daily backup. The level 1 weekly backup (Days 7 and 14), works in the same way to only copy changes since the last level 1 weekly backup. As the name implies, as data changes over time, it is *partially accumulated* at various backup levels. A new full backup will signal the beginning of a new backup cycle, but is not required to reset the size of the higher level weekly and daily backups.

These graphs were generated using backup statistics gathered from a production backup server. The data has been graphed with respect to the size of the corresponding Differential Backup to show the increased savings that PCI Backups provide over CI Backups. In this two week sample, 2 level PCI Backups were 36% smaller than the corresponding CI Backups and 58% smaller than the corresponding Differential Backup. The new PCI Backups will typically be significantly smaller in size at any backup level than currently available CI Backups. The primary reason for this is that each PCI Backup contains data changes over a smaller, fixed period of time. This “flattens” the size of the backup at each level and allows any number of backups to be taken at each level without ever increasing backup sizes. As the number of backups per each cycle increases, the savings that PCI Backups provides over CI Backups will continue to increase as well.

Alternative Use of Terminology Explained

Some vendors may refer to the term “cumulative incremental backup” when describing their level 1, Differential Backup, capability. We use the term more generally to describe the behavior of a backup at any incremental level (any level except 0) that accumulates data changes since any previous lower level backup (including level 0). Our ability to generate multiple levels of synthetic backups (not just full or differential) provides our customers with a synthetic backup technology that is far more flexible and advanced.

Some vendors may use the term “differential incremental backup” when describing their True incremental Backup capability. We believe this terminology to be confusing, especially when used in conjunction with the term differential backup, which is also considered a form of incremental backup. Additionally, some vendors may require the use of several available backup levels to implement differential incremental backup. For example, levels 4 through 9 may be needed to configure differential incremental backups on a Monday through Saturday schedule. With only three remaining non-zero backup levels, this may limit the number of ways that other lower level backups can be configured. Our True incremental Backup is implemented as a single PCI Backup level and any number of backup levels that may be configured with TiBS.

Synthetic Backups Limited by Tape Performance

TiBS is able to efficiently generate CI Backups for thousands of client volumes from a single tape using a single TeraMerge backup process. This is done by efficiently processing data in the order that it was originally written to tape. Multiple TeraMerge processes can then be run in parallel from multiple tapes using multiple tape devices for increasing speed and scale. Tape technology lacks the random access properties of disk. Tape mount delays and slow serial data access times limit the rate that data can be consolidated across multiple backup volumes during a single consolidation by a single backup process. Therefore, synthetic backup consolidation from tape has been limited to a single, previous CI Backup. Even with the limitations imposed by tape, TiBS has proved very effective at managing synthetic CI and Full backups across tens of thousands of locations, for multiple levels of backup on each dedicated backup server.

Multiple Volume Synthetic Backup in a Single Process

Our Disk Library Interface extends standard disk caching in TiBS to allow retention of any backup volume on disk in addition to or instead of traditional tape. For increased performance, TiBS has been enhanced to allow synthetic consolidation to optionally reuse data from disk instead of tape for faster synthetic backups. The process has been developed further to allow several PCI Backups on disk to be consolidated together with an optional tape backup volume in a single backup process. For example, a new full synthetic backup can now be generated by consolidating four weekly PCI Backups on disk with a single previous Full Backup volume stored on tape. Backups of this type can be performed for multiple client locations using a single TeraMerge process by processing the full backups on tape in the order that they were written previously. Multiple TeraMerge processes can then be run in parallel from multiple tapes using multiple tape devices for greater increases in speed and scale than tape based CI Backups can provide alone.

Network True incremental PCI Backup vs Network Synthetic CI Backup

The initial release of the TiBS Disk Library included multiple volume synthetic consolidation for a special type of PCI Backup known as the Network True incremental Backup. Multiple True incremental Backups could now be consolidated along with a previous lower level backup (a previous Full or CI Backup) in the generation of new synthetic backups (see *TiBS Disk Library Interface Improves Backup and Recovery Performance*). Large sites typically see about a 50% reduction in the size of their highest level daily backups when True incremental PCI Backups are employed. This savings in data size translates to faster nightly backups, reduced storage costs on the backup server (for both disk and tape), reduced wear and tear on backup hardware (especially expensive tape devices), and the capability for the same backup hardware to support an increased amount of live data.

Synthetic, Midlevel PCI Backup vs. Synthetic Midlevel CI Backup

Midlevel PCI backups offer several advantages over commonly available Midlevel CI Backups. On a TiBS backup server, since both types of backups are produced synthetically, no client or network interaction is required. Thus, the advantages of PCI Backups pertain primarily to the costs and performance issues associated with backup server hardware and related disk and tape storage.

Remove File Redundancy: By segmenting the backup time at each backup level, PCI Backups store only one copy of each file version at any given backup level. CI Backups may contain many copies of the same file version at any given level. The relative size of PCI Backups compared with CI Backups varies depending on many factors including rate of live data change, number of backups at each level per cycle (e.g. 4 weekly backups in a 28 day full cycle), and the percentage of new data versus the percentage of changed data in each backup. Over a range of sites that have implemented this new type of backup, PCI Backups have been measured to be 2 to 5 times smaller than their previously implemented CI Backup counterparts.

Increase Reliance on Disk: As disk storage costs continue to decline relative to tape, the demand for disk storage for backup and recovery continues to increase. Teradactyl maintains that a combination of disk and tape storage provides the advantages of each technology, while mitigating the risks associated with the use of only one. Since PCI backups are smaller than CI Backups, a larger window of data can be kept in near line disk storage. Disk not only accelerates backup and restore performance, it can actually reduce storage costs over the use of tape alone. PCI Backups provide a safe transition to a higher dependence and use of disk storage while maintaining the advantages and additional protections that tape still provides over the use of disk alone.

Better Use of Backup Hardware: PCI backups flatten the workload over traditional CI Backups. The reduced copying of files and increased reliance on disk accelerates backup processing. PCI Backups require less processing time, consume less storage space (both disk and tape), reduce wear and tear on expensive tape devices and require fewer tape library slots and tape mount requests than CI Backups. Multiple level synthetic backup is the cornerstone of TiBS ability to scale with data growth while controlling the cost of backup and recovery. PCI Backups provide additional efficiencies across the board for continued scale and reduced costs of this critical IT function.

Summary

Partial Cumulative Incremental Backups provide a compelling alternative to traditionally available Cumulative Incremental Backups. When used in multiple level synthetic backup strategies, PCI Backups allow sites to safely and efficiently rely more on disk than tape for backup and recovery. PCI Backups enable any number of backup volumes to be synthetically consolidated within a single TeraMerge backup process. Combined with multiple levels of synthetic backup consolidation, Teradactyl customers enjoy an increased benefit in their initial software purchase. As sites continue to grow and to rely more on disk and less on tape, Teradactyl will continue to provide solutions that allow customers to safely transition from tape or leverage the advantage of both technologies.